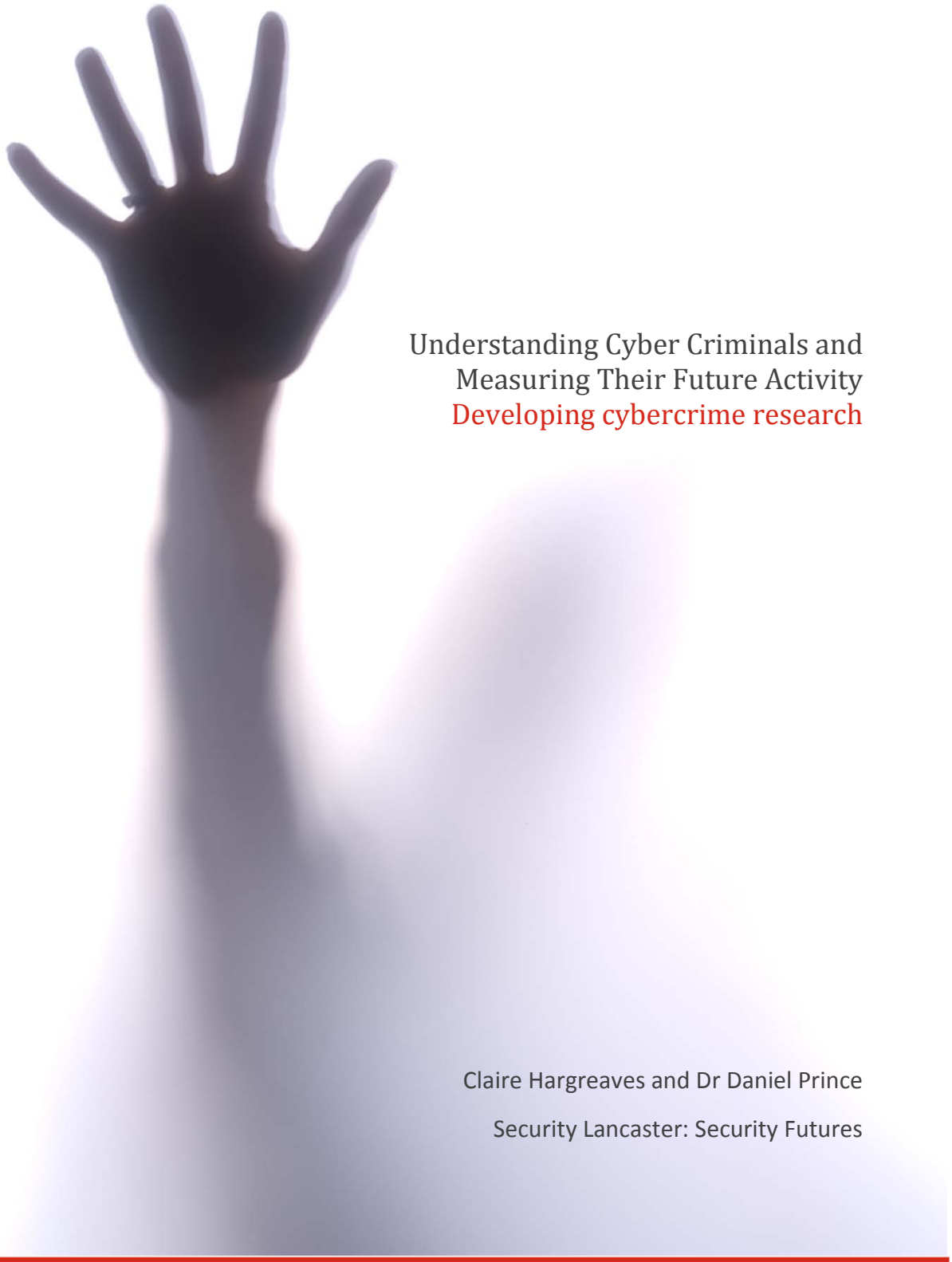
Security Lancaster | LANCASTER UNIVERSITY



Understanding Cyber Criminals and Measuring Their Future Activity

Developing cybercrime research

Claire Hargreaves and Dr Daniel Prince

Security Lancaster: Security Futures

Foreword

Contents

Prof David S Wall

Professor of Criminology and Head of the School of Applied Social Sciences. Durham University

Cybercrime is a topic in which public and governmental understandings have long been shaped by cultural constructions and the almost theatrical politics of security. In recent years, however, the Cybercrime stakes have risen with more professional and state engineered attacks, more commercial cybercrime and it being used to spread hate. As a consequence, Cybercrime has risen up the political agenda and it has come to be taken much more seriously. Our societal responses to it are now framed by Governmental Cyber-security strategies and policies, yet a simple reading of the situation suggests that we still have some way to go to improve our understanding of the actual problem.

What this interesting report by Claire Hargreaves and Dan Prince tells us is that (in the 20 years or so since I first started writing about Cybercrime) we have come a long way towards developing a real understanding of it, but there is some way to go. What I think is particularly useful about the report is the interdisciplinarity across the sciences and social sciences that Claire and Dan bring to it and also their emphasis upon the need to focus now upon the victims and offenders. It usefully sends some simple messages to the reader: we need to fully understand the role of technologies and not just make knee jerk reactions; we also need to develop standard protocols for data and also agree on mechanisms for the capture of that data; finally, there is a need to develop a more sophisticated knowledge and understanding of offenders and their victims, especially the latter who, all too often, are simply seen as a spurious data point which hides their agony.

Executive Summary

The report on the future of understanding cyber criminals and measuring their activity is created to detail the key findings from our workshop which addressed the actions required to tackle the perceived cybercrime wave.

Cybercrime is increasingly seen as a significant criminal activity by governments around the world, whether they are purely digital crimes or traditional crimes which are enhanced through the use of digital technology. Despite the anecdotally growing trend and the significant investment by governments to tackle the issues there are few publically available sources of evidence on cyber criminals. We argue that in order to be effective in tackling cybercrime a strong evidence base is required.

This report draws upon the discussions held in the workshop on defining a cybercrime and understanding the role by which the use of technology enables the criminal. We propose a classification assessment to differentiate between the two fundamental categories of cybercrime: computer enabled and computer dependent crime. We move on to explore the current state of information held, offering a data source taxonomy to facilitate the understanding of these datasets and identify the prominent features to aid data selection. During the workshop it was identified that in order to move forward in our research on cybercrime, an effort to standardise data must come into effect. The theoretical suggestions raised in this area are discussed along with how the information can facilitate research. Furthermore we detail the key points of contact at which valuable data can be collected along with current and advanced mechanisms by which information could be obtained. Following the accumulation of data and its increased quality heightened research can begin. We therefore converse proposed research on both cybercriminals and their victims.

The key findings from the workshop are outlined below.

Understand technologies role in cybercrime

We draw upon the existing literature and the discussion held during the workshop in order to define, for the purpose of this paper, a cybercrime, and to understand the role by which the use of technology enables the criminal and mediates the victim offender interaction.

Defining cybercrime was found to be a difficult task for several reasons, firstly cybercrime encompasses a broad spectrum of offences many of which can be traced back to traditional crimes, the question is then raised as to whether new definitions and laws are needed or whether amendments to existing legislation is all that is required. Cybercrime is a relatively new crime, in which government agencies, law enforcement bodies, businesses and academics have deliberated over in an effort to come to an overarching definition.

Dependent upon the organisation defining the crime some discrepancies exist, however there are two fundamental categories, computer enabled and computer dependent crime. Computer enabled crime is a traditional crime facilitated by technology whilst computer dependent crime is a crime which could not exist without new technology.

With the broad categories of computer enabled and computer dependent crime it is important to develop mechanisms to be able to judge the extent to which either should be selected. The workshop discussed utilising the following concepts:

- ***Force Amplification:*** A simple analogy here is that of physical harm with or without the use of a weapon. The level of harm that the average person can achieve with a weapon is far greater than that without. Similarly the amount of harm that can be caused with digital technology can be greatly amplified, for example, consider fraud via spam. If a fraudster had to send a letter to each target then the number of targets would be greatly diminished. The use of digital technology and communications enables the criminal to be able to interact with potential victims on a global basis.
- ***Entry Barrier:*** A significant feature of all technology is that it significantly reduces the entry barrier for people to commit a criminal activity, consider copyright theft. In this instance, the digital replication of any copyright protected information, music, film, literature, is relatively trivial given the currently available technology.

By utilising these two comparative properties crimes involving digital equipment may be compared for equivalency but also classification as computer enabled or computer dependent. This has a potentially significant advantage for the legal system in two respects. Firstly it presents the impact that technology played in the crime via a mechanism that does not rely on technical details that a lay person would find difficult to comprehend. Secondly it facilitates a comparative approach for prosecution and sentencing decisions which are again technology agnostic. This is an important principle that should be considered going forward, *Focus on the impact of the technology not on the technology itself*.

Standardise data to further our data sources

Cybercrime data is currently fragmented, this can be put down to a lack of data collection and also reoccurring arguments over the definition of what a cybercrime actually is. However, there are data sources available that can be used to help to understand the domain. The workshop sought to be surgical in its analysis of the available data sources to understand what is available and how this information may start to be combined to develop a complete and realistic understanding of the cybercriminal terrain. What became apparent in the workshop was the need for a step change in research on both cyber criminals and their victims and a more sophisticated understanding of what data is actually required to underpin appropriate analysis.

To facilitate the comprehension of these data sets we propose a classification approach that identifies the salient features of the datasets, aiding researchers to select appropriate

sources for their investigation. The proposed data taxonomy comprises five levels: originator, type, collection methodology, processing methodology and data availability. At the first level, *originator,* the researcher establishes whether the data is public or private whilst at the second level, *type,* the quantitative and qualitative levels are assessed. The scope of the target population is evaluated at the third level, *collection methodology*, and how the information has been developed should be calculated at the *process methodology* level (fourth level). At the final level, *data availability*, the level of access to the data should be considered.

Whilst the exploration of current data has its benefits, it is limited. To move forward in this space it was debated that the production and provision of a standardised data frame was needed to develop reliable and valid datasets. Delegates identified several elements mandatory in the construction of standardised cybercrime data: the data must be kept simple, wells structured, have high input standards, consistent measurements and definitions, and inclusion of basic variables. Introducing these measures will limit, at the inputting stage, mistakes such as inconsistencies and duplications. What's more, the datasets will be comparable optimizing sample size. Following the discussion on standardised data delegates debated its operationalization. It became apparent that in order for organisations to provide and maintain information, ownership of such data was key. It was suggested a structured data frame be provided to organisations in conjunction with adequate training.

Although the proposed standardised data will provide stepping stones to fill the knowledge gap, it is not without its limitations. The data will only provide a snap shot of what is occurring in the UK. Furthermore, there will be extreme difficulties in standardising data particularly when the needs of academic, government and private sectors must all be met.

Utilise mechanisms to capture data

Utilising both new and old mechanisms of data capture will develop our evidence base. The workshop set out to explore mechanisms that could be utilised to capture data on cybercriminal activity. We propose there are two vehicles of data collection that can be adapted in order to capture appropriate data. The first of these is the key points of contact through the process of criminal investigation, prosecution and sentencing. The second opportunity is periodic or asynchronous crime or impact surveys carried out by governments and businesses. Victims and offenders interact with the investigative and legislative system from the first moment a crime is reported. These interactions provide ideal opportunities to gather information regarding the involvement of technology and therefore are able to classify it as a cybercrime. This approach of, little but often, enables a mass of information to be collated as the criminal justice system process progresses, rather than in an asynchronous survey approach.

Surveys present a platform from which offenders can express their feelings, attitudes, motives and actions without fear of judicial repercussions. This method of data collection

also allows victims the opportunity to talk about crimes committed against them away from the pressures of law enforcement; victims may not feel it appropriate to report to the police or do not wish clients to know they have fallen victim to cybercrime. Furthermore, three advanced collection methods were identified: cyber specials who would be trained in both interview and cyber technology, online forums to gather more personal data and technologists who would have the ability to ascertain how equipment has been used – for better or for worse.

Broaden analysis on cyber criminals and their victims

Developing an understanding of who criminals and victims are in terms of their characteristics will help deliver appropriate interventions. Two fundamental areas were identified in the workshop as requiring extensive research: analysis of cyber criminals and investigation of victim profiles. Research on cybercrime data is minimal in comparison to the extensive analysis of traditional crime, the critical reason being the limitations of existing data. Research and its subsequent results are restricted to the quality of its data therefore advancements in cybercrime data must first be made.

Research into why individuals commit crime when others do not, and to ascertain how these individuals are different to law abiding citizens is essential if we are to tackle cyber criminality. Through the use of statistical techniques factors associated to cybercriminals can be identified allowing us to answer such questions as, do cyber criminals have common demographic characteristics? Furthermore, comparing the characteristics of online to offline offenders will help to establish whether these groups of offenders are different. The results of such analysis will assist in shaping policy in terms of detection, intervention and punishment. In addition, the development of criminal career research on cybercrime data will enhance the evidence base used by policy makers and law enforcers.

Establishing whether victims of cybercrime are a specific group of people will help to target preventative methods and resources. Through a victim information database researchers can investigate the characteristics of the victims to develop our understanding of who they are and determine if specific groups of people are more vulnerable to cyber-attacks and if so the reasons why. Following this information the common characteristics of the victims can be identified. For example, it may become apparent that the majority of victims were of the age 25 to 30, if this is the case preventative methods, such as educating them on how to stay safe online, could be targeted to this age group.

Introduction

Cybercrime is increasingly seen as a significant criminal activity by governments around the world, whether they are purely digital crimes or traditional crimes which are enhanced through the use of digital technology. Despite the anecdotally growing trend and the significant investment by governments to tackle the issues there are few publically available sources of evidence on cyber criminals.

We argue that in order to be effective in tackling the perceived cybercrime wave a strong evidence base is required. In response to this challenge a workshop was organised to explore the future of cybercriminal activity and how key stakeholders could contribute to the publically available data that should inform this evidence base. The workshop brought together experts in the field, including government agencies, legal practitioners (Pannone), and academics from multi-disciplinary areas (computer scientists, criminologists and statisticians).

The structured day-long workshop was held to explore the current situation, widen the knowledge of cyber criminals and develop innovative approaches to obtain information on cyber criminals. Importantly it provided an open environment for the stakeholders, who would normally be on differing sides in the adversarial legal process, to hold a critical, informed debate on this topic. The workshop held the following key aims:

- Understand what data already exists on cyber criminals
- To determine what information on cyber criminals is needed to bring such individuals to justice and map their criminal careers
- To explore innovative methods to capture data on cyber criminals
- To discuss the issues of storing, controlling and accessibility of a created database

This report details the key findings from that workshop that collectively the workshop participants feel could guide the policies on the response to measuring cybercriminal activity.

The key findings were:

1. Understand technologies role in cybercrime: We need to focus on the impact of technology not on the technology itself if we are to move forward in our understanding of cybercrime.
2. Standardise data to further our data sources: Cybercrime data is currently fragmented, requiring standardisation to build its reliability and validity.

3. Utilise mechanisms to capture data: Utilising both new and old mechanisms of data capture will develop our information base.
4. Broaden analysis on cyber criminals and their victims: Developing an understanding of who criminals and victims are in terms of their characteristics will help to deliver appropriate interventions.

We proceed in this report by firstly drawing upon the discussions held in the workshop on defining a cybercrime and understanding the role by which the use of technology enables the criminal. Two fundamental categories of cybercrime exist, computer enabled crime and computer dependent crime, a classification assessment of these categories has been proposed. We move on to explore the current state of information held, offering a data source taxonomy to facilitate the comprehension of these datasets and identify the prominent features to aid data selection. Following on we discuss theoretically the standardisation of data to develop reliable and valid information to facilitate research. We later review the key points of contact at which valuable data can be collected along with current and advanced mechanisms by which information could be collected. We end the report by proposing further research on both cybercriminals and their victims to advance our understanding of cybercrime.

Defining the Concepts

Cybercriminal activity has proven to be an elusive concept to define. This section draws upon the existing literature and the discussions held during the workshop in order to define, for the purposes of this paper, a cybercrime, and to understand the role by which the use of technology enables the criminal and mediates the victim offender interaction.

In his 2007/10 article, David Wall provides an approach which identifies three generations of cybercriminal activity:

- **Crimes in the machine (computer content)**
- **Crimes using machines (computer related)**
- **Crimes against the machine (computer integrity)**

Wall also goes further to identify a future where the offender victim interaction is automated by technology completely removing the need for victim selection and interaction. While this classification is useful in understanding the development and evolution of criminal activity as mediated by technology, the approach commonly used is to define computer crime as ***enabled*** (arguably the first two generations as described by Wall) and ***dependent.*** Further distinctions are also used, such as 'using technology for communication and organisation purposes of a crime', i.e. communicating online to undertake people trafficking. We use this as a starting point for the creation of definitions in the remainder of this report.

A Cyber Criminal

Important in the understanding of what it means to be a cybercriminal is the methodology by which we classify the crime as "cyber". Digital technology and mass communication networks have created new opportunities to commit crime and for criminals to interact with their victims. Cybercrime is a relatively new crime, in which government agencies, law enforcements, businesses and academics have deliberated over in an effort to come to an overarching definition. Dependent upon the organisation defining the crime some discrepancies exist, however there are two fundamental categories:

> ***Computer enabled crime***: traditional crime that is increased in its scale or reach through the use of technology. For example, phishing which attempts to acquire details (i.e. bank details) through email by purporting to be from a legitimate

organisation as opposed to an individual eluding to be the gas man to retrieve bank details.

Computer dependent crime: crime which could not exist without new technology. For example, harvesting bank account details through malware.

However, it is not clear by which metrics criminal activity is assessed in order to move it into these categories. For example, should the use of a computer to maintain a spread sheet of victim details involved in fraud be considered computer enabled? Arguably not as this information could be maintained in a paper format and the use of technology is to provide convenience to the criminal but any information gathered from the computer would be considered digital evidence. The important point here is that all computer crime has digital evidence, but not all digital evidence is part of a computer crime.

As part of the workshop, delegates were given the task of defining cybercrime. What became apparent from the discussion was the difficulty in defining such a new and ever developing crime. The problem of knowing a crime has been committed is twofold; firstly individuals need to know an act is an offence and secondly have the knowledge and skills to detect a crime has been committed. Further discussion identified inconsistencies in the attitudes and approaches used in defining cyber criminality to those used in traditional crimes. Whereas in traditional crime harm and motivation are the driving factors, methodology is seen to be the principal factor in defining cybercrime. In addition, it was felt that more attention is to be given to educating society on the harm cybercrime causes. Of those who venture into the realm of cyber criminality some are often unaware they have crossed a line, others who know their actions are wrong do not always know the level of harm they inflict.

Conversely, the importance of defining cybercrime was questioned, as even the most obscure acts of cybercrime can be traced back to traditional crimes in some way or form. Nevertheless, it was deemed that there are unique offences which should be categorised as separate to the ones within the general criminal spectrum, such as the modification of computer material without authority to do so.

To be classed as dependent or enabled a crime must have some unique features to it that involve the use of technology to move it beyond a traditional crime. The dependent classification has a clear delineation that it is only possible with the use of technology, for example illegally selling guns would not be a crime if gun technology did not exist. A significantly more difficult judgement is what causes a crime to be classified as enabled. Here an assessment of the impact the technology has on the commission of a crime and the way the offender/victim interaction is mediated is discussed.

Impact of Technology

With the broad categories of computer enabled and computer dependent crime it is important to develop mechanisms to be able to judge the extent to which either should be

selected. One such selection mechanism is the transformation or elimination test (Wall, 2007), where the extent of the crime is assessed when technology mediation is removed from the criminal act or whether it could be substituted with other non-digital methods or approaches. This assessment goes someway to identify whether the use of digital technology in a crime should cause it to be identified as a cybercrime. However, we would argue that this presents a rather binary measure of the use of technology in the commission of a crime. As an alternative the workshop participants discussed whether an assessment of the impact that technology had on the commissioning of the crime rather than having a technologically detailed focus provides a more nuanced, graduated assessment process. Further, by having technologically independent assessment mechanisms would avoid having to repeatedly update classifications based on rapidly evolving technology. The workshop discussed utilising the following concepts:

- ***Force Amplification:*** A simple analogy here is that of physical harm with or without the use of a weapon. The level of harm that the average person can achieve with a weapon is far greater than that without. Similarly the amount of harm that can be caused with digital technology can be greatly amplified, for example, consider fraud via spam. If a fraudster had to send a letter to each target then the number of targets would be greatly diminished. The use of digital technology and communications enables the criminal to be able to interact with potential victims on a global basis.
- ***Entry Barrier:*** A significant feature of all technology is that it significantly reduces the entry barrier for people to commit a criminal activity, consider copyright theft. In this instance, the digital replication of any copyright protected information, music, film, literature, is relatively trivial given the currently available technology.

By utilising these two comparative properties crimes involving digital equipment may be compared for equivalency but also classification as computer enabled or computer dependent. This has a potentially significant advantage for the legal system in two respects. Firstly it presents the impact that technology played in the crime via a mechanism that does not rely on technical details that a lay person would find difficult to comprehend. Secondly it facilitates a comparative approach for prosecution and sentencing decisions which are again technology agnostic. This is an important principle that should be considered going forward, *Focus on the impact of the technology not on the technology itself*.

As the field of technology within the legal system is driven by technologies there is a natural tendency to focus on the details of the technology rather than the impact that it has on the facilitation of the crime. While there is certainly a need to obtain accurate details of technological use in commissioning a criminal activity in a forensically sound manner, it is vital not to lose sight of the impact this technology had on the criminal act. For example, a death threat can be sent by post or by email, if it is by email it should not make it a computer enabled crime. Given that this was only an initial exploration of qualifying the impact of technology on the criminal act, we do not claim that the criterion above is

completely appropriate. Further work needs to be completed in order to identify or validate appropriate criterion and a suitable, rigorous and repeatable methodology for classification against these criteria needs to be developed. However, the possibility of being technology agnostic in the assessment of technology in the criminal act as shown in Figure 1 is intriguing.

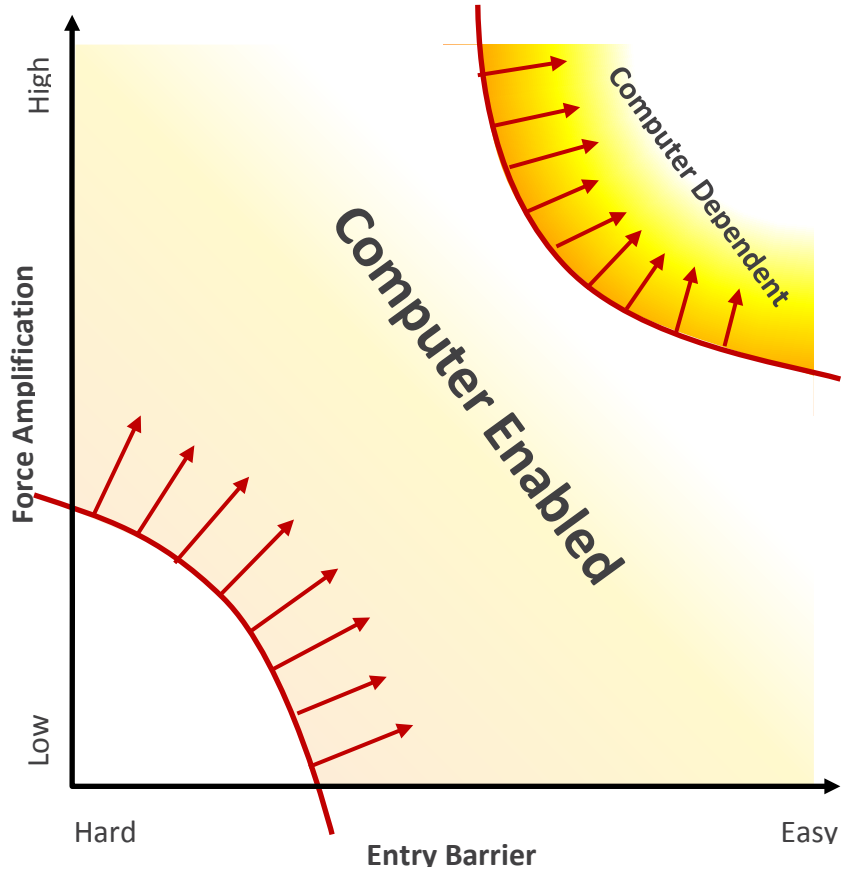


Figure 1: Classification of computer enabled and computer dependent crime

This begs the question of whether the use of technology to commit a crime should in fact be taken into account in the sentencing of a convicted criminal, in the same way the use of other types of technology, cars, weapons etc., are utilised in a criminal act may also impact sentencing. However, this discussion is beyond the scope of the workshop and this report.

Summary

Defining cybercrime was found to be a difficult task for several reasons, firstly cybercrime encompasses a broad spectrum of offences many of which can be traced back to traditional crimes, the question is then raised as to whether new definitions and legislations are needed or whether amendments to existing laws is all that is required. Secondly, in the various forms of cybercrime a technological skill set is often required thus a limited understanding of such skills widens the barrier to understanding cybercrime. A better understanding of the role by which the use of technology enables the criminal and mediates the victim offender interaction is needed. In addition, the criterion used such as, harm and

motivation, in the current definition of cybercrime do not match those used in traditional crimes. Could rectifying this formula further the effectiveness of the criminal process and also aid society in its understanding of when a cybercrime has been committed and the implications of such acts?

However, although there are issues in the clarity and understanding of cybercrime, the criminal justice system has still been able to prosecute these offenders (however small). The current definition of cybercrime has two fundamental categories: computer enabled and computer dependent. Is it then that supplementary guidelines or classifications on the impact of technology to establish if a crime is computer enabled or dependent, all that is needed?

Cybercrime Data

The picture of cybercrime is currently fragmented and incomplete. This can be put down to a lack of data collection and also reoccurring arguments over the definition of what a cybercrime actually is. However, there are data sources available that can be used to help to understand the domain.

The discourse surrounding these data provides a conflicting environment where public and private sector interests collide, media outlets generate hype in public opinion to create newsworthy stories and the potential of cyber-attacks is often misrepresented as fact. This environment has resulted in the UK government prioritising cyber security as a Tier 1 national security threat (injecting £650M in cash support) and cybercrime remediation as a key component in its strategy to promote the UK in a digital economy. The workshop sought to be surgical in its analysis of the available data sources to understand what is available and how this information may start to be combined to develop a complete and realistic understanding of the cybercriminal terrain. What became apparent in the workshop was the need for a step change in research on both cyber criminals and their victims and a more sophisticated understanding of what data is actually required to underpin appropriate analysis.

This section attempts to provide mechanisms to classify and evaluate the quality of available data. It will examine the advantages and disadvantages of existing data (i.e. inconsistent measurements, different definitions), which will be followed by a discussion of a standardised data set on cyber criminals and their victims in which criminal career and victim profiling can be implemented.

Data Source Taxonomy

Data sources are the lifeblood of understanding criminal activity, methodology and victim profiles. There are a range of data sources available for researchers to use from a variety of sources such as commercial (e.g. Internet service providers, third party payment authorisation companies and vendors) and government departments (e.g. National Fraud Intelligence Bureau, Action Fraud and Crime Survey of England and Wales). However, the types of data source, the way it is presented and gathered all have a bearing on the usefulness of the data in developing an understanding of these areas. To facilitate the comprehension of these data sets we propose a classification approach that identifies the salient features of the datasets which will aid researchers to select appropriate sources for their investigation.

A fundamental data feature used for classification of sources is whether that data is qualitative or quantitative. In the context of data collection in computer incident reporting from sources such as Symantec, hard quantitative figures of malware infection are easy to collect from the devices that have the providers software installed. Similarly the financial impact of a criminal activity on a victim can be estimated and quantified. Quantitative data is useful in developing statistical numerical approaches to understanding criminal activity however it lacks the ability to capture important aspects of the criminal operation, motivation and impact. In this instance qualitative data enables richer datasets to be collected however, these data are more open to the subjective interpretation of the research in terms of reporting the results. In these instances a well-defined methodological approach to data capture and analysis is required to develop confidence in the produced dataset.

Further differentiation can be found in the owner of the data source. Common sources of the data regarding the impact of cyber-attacks on the business community are collected by business groups themselves to explore specific populations. A further observation regarding these private sector reports is that they are also primarily technology orientated, focusing on the information that can be gathered from their services and products. Even surveys querying the community focus on the impact of technology use, consumption and violation. In contrast public data sources in this area have a crime focused approach, collecting details on the criminal act, the methodology and the outcome of the crime on the victim. It was also identified during the workshop that judicial reporting sources may often miss or mask a cyber component due to the precedent to focus on prosecuting crimes that would render the maximum impact on the suspect or has the most significant changes of success. A further accusation that has been levelled at private sector data sources is that they are typically structured in such a way as to provide confirmation bias in support of the business' commercial position. This accusation is compounded as private sector reports are unlikely to be accompanied with the associated raw data sets and details of methodological approaches to collect and process the data resulting in a lack of transparency in the reported outcomes. Similar accusations can be levelled at public data in terms of support for the government of day, however, it is much more likely that public data is made available in its raw form to be verified by third parties.

Another fundamental feature which defines the data is the method of data collection with a key aspect of this being the target population. The scope of the target population has a bearing on the range of responses collated. For example, some data sets often focus on key target groups such as business owners, technical personnel and so on, potentially limiting the range of responses. This is in contrast to assessments where great care is taken to obtain a representative group or where the data captured is from events beyond the control of the surveyor, for example the Crime Survey for England and Wales (CSEW) asks people aged 16 and over living in households in England and Wales about their experience of crime in the last 12 months. Information on both household and personal crime are obtained

through this survey. Beyond a discussion on the target population is the actual methodology used and the availability of information regarding the approach. Information regarding the approach used, double blind surveys, incentivised surveys, data gathering from computer programmes, all have a bearing on the reliability and confidence in the data. This is particularly important where novel methodologies are required to capture the data necessary to understand the unique features of cybercriminal activity.

In addition to how the data is collected the availability of information on the processing methodology used is vital. This is of specific importance in the case of qualitative data sets and of other instances where there is a significant level of subjectivity in the interpretation of the data. For example, consider text based answers to the methods used by a cybercriminal. These data are clearly qualitative and therefore a subjective classification is needed in order to aggregate the data for reporting purposes. Here, having a rigorous, repeatable method for analysing data that can be performed by a third party and generates equivalent results is vital. While quantitative methods are more standard having a clear understanding of how the reported data were derived adds confidence to the quality of the reported interpretation and allows for repetition of analysis and verification of results.

A final category that we feel should be considered is the level of access that is available to the data underpinning the research. A common practice, especially in the private sector, is to only publish the interpreted findings, i.e. the aggregated data. There are various motivations for this approach which are largely to do with data protection compliance. Anonymising such data, whether qualitative or quantitative, is complicated and an expensive task with potentially significant fines if identifying data is made available. However, the lack of raw data can undermine confidence in the published results as there is no way to verify the reported findings. The challenge of making this type of data available need to be overcome if we are to be able to look at third party sources, such as credit card companies, who may hold valuable information on cybercriminal activity. The problem is further compounded in the legal profession which could potentially hold a rich source of information on these types of crime, but would be bound by a stringent client-advocate confidentiality policy. The extreme alternative is to publish the full open data set for analysis by third parties, this openness should be coupled with information on the analysis/processing methodology in order to replicate the reported findings.

A summary of these categories is given in Figure 2. For brevity and read ability, the complete taxonomical tree is not presented and it is assumed that each layer has a set of classification nodes connected to the parent.

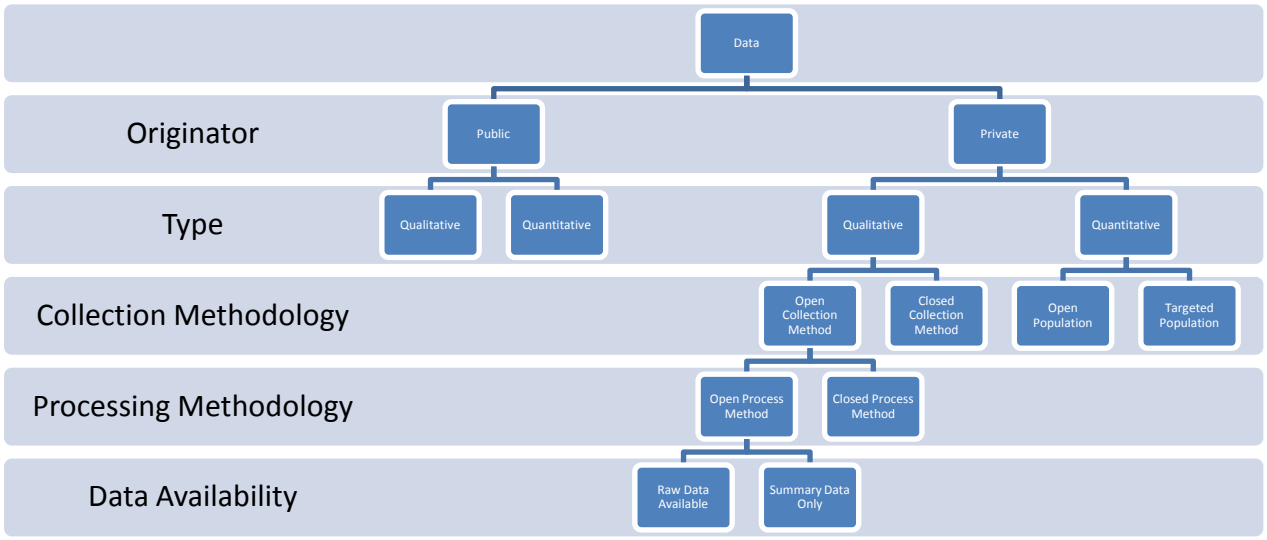


Figure 2: Classification Taxonomy of data sources

Blue print of cybercrime data

Current data on both cyber criminals and their victims come in many shapes and sizes (discussed above), presenting difficulties in the collation and analysis of data. What became very apparent from the workshop was a need for the standardisation of cybercrime data. The workshop enabled informative discussion on the required improvements of constructed data and the framework needed to develop its reliability and validity. Here we discuss theoretically the creation and workings of standardised data. The discussion encompassed three elements: how data should be standardised, the operationalization of standardised data and the limitations of such constructed records. It was felt that each organisation (government, private sectors and academics) should run and maintain their own standardised data, increasing the collection of generic information with the addition of data specific to the organisation, which can be shared and analysed collectively.

Delegates identified several elements mandatory in the construction of standardised cybercrime data. Primarily the data must be kept simple, well-structured with high input standards - an element lacking in many databases. The dataset should be structured clearly that enables easy navigation, for instance, the information each variable holds. If the data and usability is not simple mistakes are imminent as the clarity of use and navigation diminishes. Having incoherent variable names and vague instructions of what data is to be inputted leads to inconsistencies and duplicates within variables causing problems when searching data; limiting inputters' discrepancies when keying data can reduce this problem. A method suggested is the use of drop down menus for selection. For example, in a variable identifying the birth country of an offender, having no drop down menu may result in an inputter misspelling the country or writing the country with a capital letter the first time but not the preceding times.

Furthermore, measurements and definitions must be clearly stated to ensure they remain consistent across datasets. For instance, datasets with varying definitions of cybercrime renders them incomparable. Moreover, individuals may be recorded multiple times under different aliases; the datasets must be able to manage multiple identities to ensure the same person is not counted more than once. The workshop identified a further key requirement of standardised data, the inclusion of several basic variables. Two reasons exist, one for comparability purposes and the other to obtain optimal sample sizes on both offenders and victims' basic information.

Following the discussion on standardised data delegates debated its operationalization. It became apparent that in order for organisations to provide and maintain information, ownership of such data was key. It was suggested a structured data frame be provided to organisations in conjunction with adequate training. The data frame would include mandatory variables for data collection plus the ability for owners to add additional factors. Due to the sensitivity of the data it was deemed compulsory to implement access rights dependent upon status (i.e. student, academic, private, public, government). Owing to the various organisations that may be undertaking operations or investigations based on the data held in such a system there exists a requirement for key information sets to be *flag-able*, that is to be able to set public notification that other services are utilising the data, or to set private requests for notification when services review key data pertinent to their investigation. Such capabilities are intended to help prevent wasted manpower or impact between multiple, simultaneous investigations. It was believed that in order to capitalize on the standardised data both in terms of criminal justice and research, an umbrella search engine will be required. The search engine would process enquires on content contained throughout the datasets, feeding back the information whilst abiding by access rights. In addition, ownership information would be provided to assist request for access.

Although the proposed standardised data will provide stepping stones to fill the knowledge gap, it is not without its limitations. The data will only provide a snap shot of what is occurring in the UK. Furthermore, there will be extreme difficulties in standardising data particularly when the needs of academic, government and private sectors must all be met.

Summary

The production of a five level data taxonomy, encompassing: originator, type, collection methodology, processing methodology and data availability, will provide researchers with the tools to classify and evaluate the quality of existing data. In doing so, investigators can select the appropriate sources for their investigation increasing the quality of results on which policy can be formulated. Without the mechanisms to identify the potential limitations of data, the risk of unaccounted bias increases greatly. However, to push forward in this space, new standardised data collections need to be implemented to develop reliable and valid datasets whilst facilitating optimal sample sizes and comparable datasets. Continued research on both existing data and the production of new standardised data will provide a clearer picture of cybercrime.

Mechanisms of Data Capture

The workshop set out to explore mechanisms that could be utilised to capture data on cybercriminal activity using the types of innovative methods that digital technologies are capable of providing. However, the attendees quickly identified that a fundamental issue with current data collection approaches was that basic information, such as email addresses, were not being collected.

This discussion broadened out into the key points of contact at which valuable, basic data could be collected and finally the mechanisms by which this basic information could be collected and how it could be extended to more advanced and nuanced data collection techniques. There are two significant barriers to the collection of information on cybercriminal activity. The first, and possibly the most significant barrier is the low threshold of the majority of cybercrime. The use of digital, mass communication technology enables the aggregation of numerous low level crimes that when aggregated comprise a significant overall prize to the criminal. As Wall notes, these crimes often fail to reach a threshold set by the investigative bodies or the victims that determines whether they are worth pursuing, which he describes as the *de minimis* trap (*de minimis non curat lex*: The Law does not concern itself with trifles). As a result the individual crimes go under reported. The second barrier is the distributed nature of attacks on such large numbers of victims making it very difficult to co-ordinate evidence gathering which would lead to aggregating cases together. This second barrier is further complicated by the difficulty of attribution of a digital crime to the offender as they have the ability to maintain and generate a plurality of identities and to attack victims from multiple international locations. These barriers present a unique problem for law enforcement as without accurate data collection accompanied by suitable correlation mechanisms on large scale organised[1] cybercriminal activity, the investigative system may never become aware that a criminal activity has reached the *de minimis* assessment threshold.

While fundamental to the investigative and enforcement process, the analysis of cybercrime also requires very similar data, statistical processes and correlation techniques to understand the cybercriminal activity in terms of offender motivation, offender actions, victim profile and criminal trends. We therefore propose there are two vehicles of data collection that can be adapted in order to capture appropriate data. The first of these is the key points of contact through the process of criminal investigation, prosecution and

[1] By organised, we do not make reference to organised crime, but rather a wide spread criminal activity that targets a significant number of victims, perpetrated by either an individual or small group of individuals. This does not preclude the offenders being part of a large scale structured criminal enterprise.

sentencing. The second opportunity is periodic or asynchronous crime or impact surveys carried out by governments and businesses. Of these the former is possibly the most important data capture source to get right as it will inevitably form a primary source while the information gathered by the latter process is often tempered by the lens of societal attitudes at the time the survey is conducted.

Key points of contact for information collection

Victims and offenders interact with the investigative and legislative system from the first moment a crime is reported. These interactions provide ideal opportunities to gather information regarding the involvement of technology and therefore are able to classify it as a cybercrime. This approach of, little but often, enables a mass of information to be collated as the criminal justice system process progresses, rather than in an asynchronous survey approach. In order to map the key points of contact where information could be collected on both the offender and the victim a basic model of the investigative and prosecution process in the UK has been provided in Figure 3.

Entry into the System

A key point of data collection is where both the offender and the victim enter the system. During the workshop discussion it was highlighted that simply collecting an offender or victims email address(es) would greatly aid in the investigative process of the specific case and also provide intelligence for correlation activities where fake or fraudulent digital identifies are used in investigations of digital locations such as bulletin boards. For the suspect key data could be gathered during the initial interview or if the suspect is ultimately arrested during questioning by the custody desk officer. Victims arguably provide a richer primary source of information regarding the crime they are a victim of. From the data collected at this point hypothesis can be tested through statistical techniques. For instance, recording the age at offence will allow the hypothesis *cybercriminals begin offending as a juvenile* or *the mean age of conviction for females is the same as that for males* to be tested.

Prosecution and Pre-trial Services

During the development of the case by the state prosecution service, decisions are made as to which crime to charge the offender with (if there are numerous possibilities) based on criteria such as the strength of the evidence in respect of the offence and the maximum punishment that each offence can yield. The result is to provide the strongest possible prosecution case which would result in the maximum penalty. These decisions may mask the involvement of a digital, technological component to the case which may not surface in later information regarding the prosecution. As a result these key decisions need to be recorded and made available so that they can be analysed in order to answer questions such as whether cybercrimes are not being prosecuted because of the complexity of the case and therefore prosecutors are reticent to proceed. Multi-level modelling, which expresses how the response variable depends on, or is explained by the explanatory variables, can be

conducted on information gathered here. The outcome of judicial proceedings can be assessed taking into account several hierarchical levels: offender, offence type, court, county. From this analysis we can see the effect of these hierarchical levels on the prosecution of cybercrimes.

It was also pointed out during the workshop that as part of the development of the case to present to court, the prosecution will also seek external legal opinion with regard to the digital nature of the crime. The information provided by the third party then informs the prosecution services decision to proceed. This action should be captured if a decision capturing process as described above were to be implemented. However, it does provide the possibility for another source of survey data, approaching respected third parties who prepare such opinions and requesting a release of data regarding the number of requests and key data points regarding the opinion.

Adjudication

The adjudication process naturally provides information on the outcome of the prosecution such as acquittal. However, as Figure 3 demonstrates there are a number of other key decisions regarding where a trial is required to take place and also why that decision was made. A key decision as to whether a magistrates' court undertakes the trial could be the involvement of digital technology in the commission of the crime.

Sentencing and Sanctions

As highlighted in the previous section, there exists the possibility that the use of digital technology in the commission of a crime may impact on the sentencing of the convicted criminal. In such cases, it would be beneficial to understand the role in sentencing that digital technology played in mediating, either as aggravating or reducing factors, the decision, and also the justification for such decisions. Such information could form the basis for revising sentencing guidelines which are currently limited in their scope with regard to guidance on the impact that technology has an aggravating of reducing factor. In addition, the collection of sentencing information such as conviction dates will enable the implementation of advanced statistical analysis including survival analysis (also known as event history analysis). Such statistics will test the hypothesis *cybercriminals do not desist from offending*.

Corrections

While not necessarily an active data point collection, it does provide an ideal point for offender surveys in order to understand motivations and techniques. Further, although sentencing provides data on how long a criminal is intended to stay imprisoned, the prison system provisions for early release of prisoners via, amnesty, commutation of sentence or parole. Information regarding these outcomes completes the picture of the cybercriminal lifecycle and may provide further insights into the evolution of the cybercriminal, for example in terms of reoffending rates.

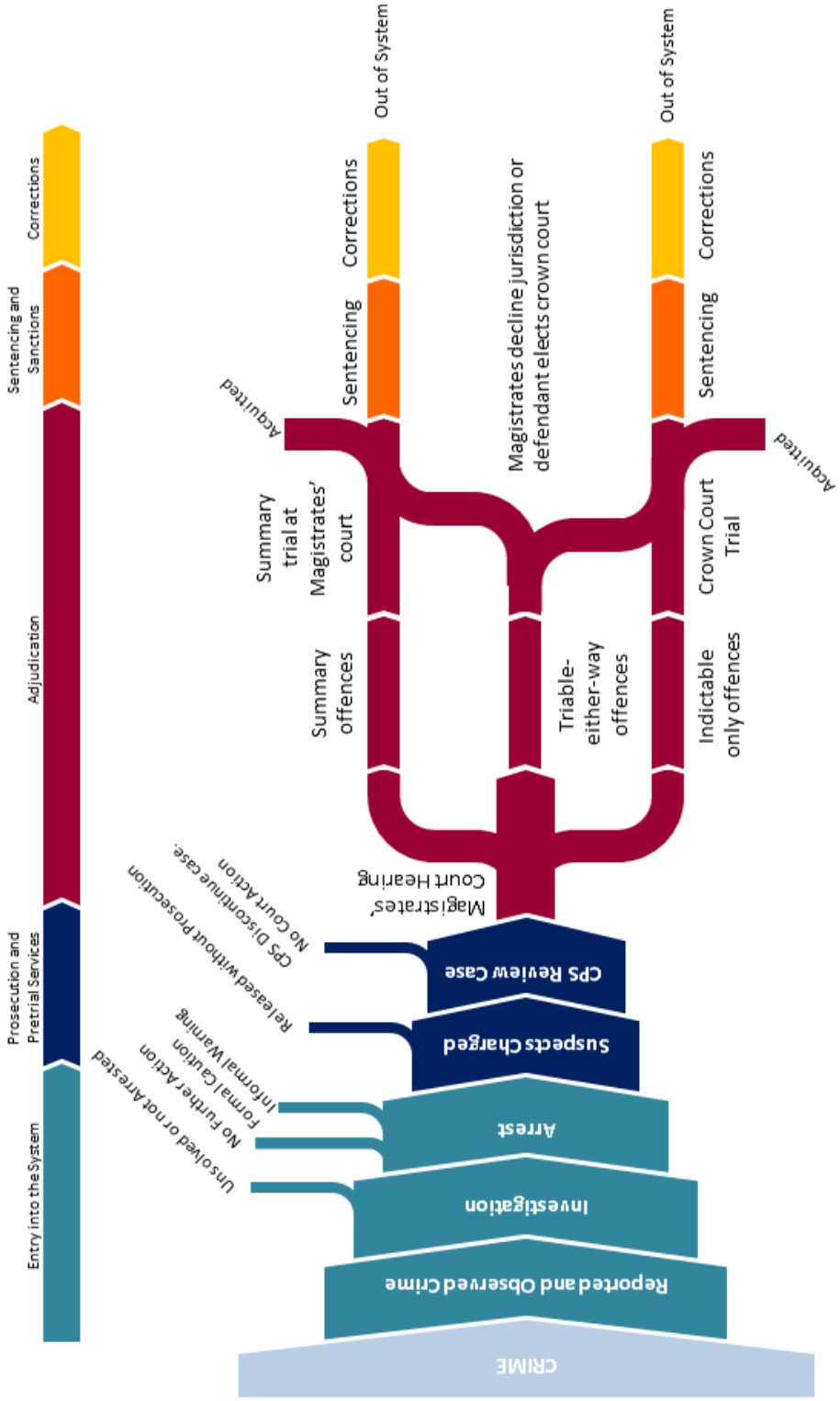


Figure 3: The criminal justice process

What is the basic information that should be collected

Considering the existing data collection points recognized above the workshop identified the basic information to be collected. Table 1 indicates the collection points at which the identified information can be collected. The table also declares which category the data comes under (i.e. demographic, socio-demographic, criminal history, attitude, and motivation), and an indicator to whether the information can be collected on offenders and victims.

All aspects of the criminal justice system provide rich sources of information that can be used to understand cybercriminal activities. However, it is important to realise that the purpose of the system is to bring offenders to justice not to generate information for statistical research. Therefore, any undertaking the criminal justice system makes in collecting data must have minimal impact on the personnel collecting the data otherwise it will be seen as an additional burden that may lead to inaccurate or incomplete data being collected. It is also vital to realise that the type of information that enables the classification and understanding of cybercrimes is sparsely distributed over several potential collection mechanisms and therefore new approaches must be adopted in order to provide a complete picture. However, it is not known whether cybercrime is substantial enough to warrant the expenditure it would take to implement such data collection, correlation and aggregation capabilities within the criminal justice system without appropriate evidence, creating a *chicken and egg* paradox. This paradox may be remediated via the use of reliable sources collating data by other approaches.

Table 1: Basic data collection points

Category	Information	Offender Information	Victim Information	Data Collection Point				
				Entry into System	Prosecution and pre-trial services	Adjudication	Sentencing and sanctions	Corrections
Demographic	Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Date of birth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Address	Yes	Yes	Yes				Yes
	Email address	Yes	Yes	Yes				
	Internet footprint	Yes	Yes	Yes				
Criminal history	Arrests (incl. previous)	Yes		Yes				
	Convictions (incl. previous)	Yes		Yes			Yes	
	Offence at arrest	Yes		Yes				
	Offence at trial	Yes			Yes			
	Offence at conviction	Yes					Yes	
	Trial proceedings	Yes			Yes	Yes	Yes	Yes
Attitude	Do they think they have committed a crime?	Yes		Yes				
	Attitude to law enforcement	Yes	Yes	Yes				
	Self-protection – do they put up barriers to prevent detection	Yes		Yes				
	What is your attitude to cybercrime?		Yes	Yes				
	Risk – did you think you were at risk to cybercrime?		Yes	Yes				
Motivation	Motivation to commit the crime	Yes		Yes				
	Motivation to move to online crime	Yes		Yes				
	Intention	Yes		Yes				
Technology	Do you use the internet		Yes	Yes				
	How often do you use the internet?		Yes	Yes				
Victim	Knowledge of crime – how did you know you were the victim of cybercrime?		Yes	Yes				
	Type of cybercrime – do you know the type of cybercrime?		Yes	Yes				

Adapting Existing Surveys and Research

While the criminal justice systems adapts its data collection processes to act as a primary information source on cybercriminal activity, alternative data sources need to be identified to help bridge the void in the currently available data. As highlighted previously many of the key companies operating in the cyber security sector provide reports that detail their findings on numerous aspects of cyber security from breaches to financial impact. However, care must be taken in the consumption of such data as it is not always reported in a way to further the agenda of the reporting agent. There is also typically a slant toward a business audience in order to engage the business community with the issues surrounding the economic impact of cyber security on business prosperity. This leaves data sets that are produced in the public interest, such as the Crime Survey for England and Wales, Action Fraud (who for computer dependent crimes and online fraud distinguishes between online and offline crimes), or academic research data sets. In the 2010/11 crime in England and Wales report, the UK national statistician highlighted that cybercrime information "was not covered well by main statistics or alternative sources." Information sources such as the CSEW and Action Fraud will be vital going forward as a first stage in gathering reliable data on cybercrime. However the question still remains as to what information would be useful to collect in such a survey.

Surveys present a platform from which offenders can express their feelings, attitudes, motives and actions without fear of judicial repercussions. This method of data collection also allows victims the opportunity to talk about crimes committed against them away from the pressures of law enforcement; victims may not feel it appropriate to report to the police or do not wish clients to know they have fallen victim to cybercrime. The following pieces of information are a number of examples of the data that could be theoretically gained from surveys (on offenders and victims) to help us understand the reasons behind cybercrime:

Offender

- Whether the think they have committed a cybercrime
- Their attitude to law enforcement
- Who they believe their biggest threat is in relation to their criminal acts
- Do they self-protect, i.e. put up barriers to prevent themselves being caught
- How the criminal act was conducted
- How they become involved
- Motivation to commit the act

Victim

- Did they think they were at risk of a cyber attack
- Do they feel they should protect themselves on the internet
- Use of the internet
- Brands and equipment used in daily life

- How they knew they were victim of cybercrime

Self-report studies can help to obtain information unknown to authorities. This mode of collection is extremely important when building a picture of crime as it provides information which would not necessarily have been obtained through official statistics.

Universities also present a fundamental source of data. They have the ability to undertake research on a more tightly focused area in order to investigate particular aspects of cybercriminal activity[2]. Whereas national bodies such as the office for National Statistics has to be more focused on obtaining information on broader activities and trends. The combination of these bodies provides an ideal platform to undertake research to understand cybercriminal activities at multiple levels. However, it would be prudent to ensure that such research activities are considered as part of a high level research strategy that is prioritised to help deliver answers to the pertinent questions around cybercrime.

Advanced data collection: Where do we need to be?

Following on from the workshops discussion on existing data collection points, the delegates deliberated how the basic mechanisms of data collection could be extended to more advanced and nuanced data collection techniques. Three principal collection methods emerged: development of cyber specials, online forums and technologists. The development of 'cyber specials' who would be trained in both interview and cyber technology would help to bridge the technology gap between law enforcers and offenders. Their training would give them the knowledge to understand offenders actions and in some cases victim technology which may have resulted in them falling target to cybercrime. Such information would help us to understand the methods used in these criminal acts and provide policy with the necessary information to implement preventative methods. Previously we discussed the current collection points (Figure 3) at which data can be collected, considering these points it would be most beneficial for cyber specials to help in the collection of data on entry into the system. Their training in interviews and cyber technology will give them the knowledge to interview offenders on entry into the system, collecting information on the technology used, methods implemented, self-protection methods, how they became involved, detection, motivation and intention.

Online forums was the second principal to emerge in which law enforcement, government agents, and researchers can talk to cyber offenders and potential cyber criminals to build a wealth of information. Research of cybercriminals using online forums has already begun; Holt, 2010; Holt *et al.,* 2008; Holt *et al.,* 2012. Data such as crimes committed, techniques used, intentions and motivations could be collected along with their demographic and socio-demographic information. Collecting this array of information would facilitate the statistical analysis to establish significant factors related to offenders. On discussion of data collection

[2] Naturally universities may also undertake broad based data collection activities.

within online forums it was felt that trust between collectors and offenders may limit data assembly, therefore significant knowledge, research and effort will be required to gain the trust of the online forum. Furthermore, it was deemed necessary that those who would enter the online forums would need to be skilled enough (e.g. cyber specials) to understand the technological methods discussed and norms of the *cyber world*.

Technologists were seen as the third principal collection method. Their knowledge would enable them to decipher if the equipment has been used for their original intentions or for criminal actions. It is seen that the technologists would in some instances be able to remotely identify if equipment is being ill-used and also investigate held equipment. Information on the equipment being used to carry out these attacks and their procedures will aid preventative methods and detection. Correlations can then be made between the equipment used and type of cybercrime.

Summary

To increase both the quantity and quality of our data we must both utilise existing data collection points whilst developing further advanced techniques to capture those untouched areas of information. It became apparent in the workshop discussion that basic information still needed to be collected in which current collection points could be exploited. The criminal justice process provides an abundance of contact points to which basic information can be collected. Five stages of the process exist: entry into the system, prosecution and pre-trial services, adjudication, sentencing and sanctions, and corrections. Exploiting each collection point will aid the development of the investigative and enforcement process whilst benefiting the analysis of cybercrime. In addition to the criminal justice process existing survey and research provide alternative data sources. Surveys present a platform from which information can be collected that is quite often missed within the criminal justice process such as, feelings, attitudes, motives and actions. Furthermore, three advanced collection methods were identified: cyber specials who would be trained in both interview and cyber technology, online forums to gather more personal data and technologists who would have the ability to ascertain how equipment has been used – for better or for worse.

Cyber Criminals and Victims

A recurring theme throughout the workshop is the need for collaborative work in this space. It was felt that in order to effectively research such a new and unique area, collaboration between government, industry and academics is needed. Whilst the government provides high level research, the time and resources available to them are limited; collaborating with academics will enable the expansion of time and resources. In addition, academics will provide the trust needed for direct communication with offenders to gather extensive information.

Two fundamental areas were identified in the workshop as requiring extensive research:

- Analysis of cyber criminals
- Investigation of victim profiles

This section discusses the proposed research on both cybercriminals and their victims. Developing an understanding of who criminals and victims are in terms of their characteristics in addition to existing data sources could help to prevent future crime, deliver appropriate interventions and safeguard the public. Research on cybercrime data is minimal in comparison to the extensive analysis of traditional crime, the critical reason being the limitations of existing data (discussed in 'Cybercrime Data'). Research and its subsequent results are restricted to the quality of its data therefore advancements in cybercrime data must first be made.

Analysis of cyber criminals

Research into why individuals commit crime when others do not, and to ascertain how these individuals are different to law abiding citizens is carried out worldwide. Within the workshop research into the profile and reasons of cyber criminality were discussed theoretically at length with numerous questions raised;

Do cyber offenders have common demographic characteristics, for example, are they a particular age? Do only males commit cyber offences? Where do cyber criminals live, rural, urban? Do they reside in the same country? Do those who specialise in a particular form of cybercrime live in a specific area? Or is it that these individuals have similar attitudes, for instance, not believing they have committed a crime or oppose the law? Or do these attackers have similar motives or a similar skill set.

Answering the above questions and comparing the characteristics of online to offline offenders will help to establish whether these groups of offenders are different. The results of such analysis could help in the detection, intervention and punishment of such criminals. As already noted cybercrime is a unique crime, requiring a new set of skills not previously needed in traditional crime.

Criminal career research, which is fundamental to criminological study, will enhance our evidence base on cybercrime.

> "The construct of a 'criminal career' is a powerful approach for accumulating rich knowledge about offenders and using that knowledge for developing rational policies for dealing with crime… The basic thrust of criminal career research is to look at the characteristics of individual offenders and to use that information for dealing with a particular individual, but also to look at aggregates of offenders and their collective characteristics for guidance on how criminal justice policies can best respond to their offending patterns." (MacLeod *et al.* 2012).

Following the formation of longitudinal data on cyber criminals, which is a key component of criminal career research, we can begin to make inference on cyber offenders' criminal careers. Such research considers the impact of age or stage of career, on factors that influence criminal behaviour and attempts to explain motivations for starting, continuing and stopping offending. Several features of an offenders' criminal career can be studied; onset, duration, frequency, seriousness, escalation, desistance, prevalence, specialisation and recidivism (see Farrington, 1992; MacLeod *et al.* 2012; Soothill *et al.,* 2009).

Examining just a few features of an offender's career can provide a wealth of knowledge, to illustrate; onset examines the start of an individual's criminal career, for instance, establishing the age at which someone begins offending, or the entry offence into their criminal career. Frequency looks at the occurrence of offences, such as, the number of cybercrimes reported, prosecuted and convicted. Specialisation investigates how far offenders focus on certain types of offending, for example, do cyber criminals commit only cyber offences, which can be extended to whether they commit only one type of cyber offence or multiple (i.e. phishing, fraud, espionage). Moreover, prevalence looks at the proportion of the population who are committing offences, for instance the proportion of cyber criminals in the UK.

Whilst the results will provide valuable information, measuring the varying features of a criminal career is extremely difficult laying host to a number of limitations, the principal being the samples true representation. The quality of the data and the information stored (discussed in 'Cybercrime Data) restricts the analysis and area of the criminal career which can be researched. For instance, official data only reveals a limited number of the total amount of crimes committed. This is due to not all crimes being reported, not all reported crimes being recorded and not all offenders being apprehended. Furthermore, results

obtained through the analysis of official crime data will measure the criminal process and changes within this process along with criminal activity (Francis *et al.* 2004).

Investigating victim profiles

The second fundamental research area identified in the workshop is the investigation of victim profiles, in which delegates identified three key questions:

- Are victims of cybercrime a specific group of people?
- Are there reasons why these individuals are becoming victims of cybercrime?
- Does the victim base change by type of cybercrime?

Establishing whether victims of cybercrime are a specific group of people will help to target preventative methods and resources. The government aims to make the public more aware of the harm caused by cybercrime and the preventative steps which can be taken to help safeguard from such attacks. For example, the website www.getsafeonline.org has been published providing informative information to stay safe online.

Through a victim information database (discussed in 'Cybercrime Data') researchers can investigate the characteristics of the victims to develop our understanding of who they are and determine if specific groups of people are more vulnerable to cyber-attacks and the reasons why. The first step is to identify and explore the demographics (i.e. age, gender, location), socio-demographics (i.e. education, employment, income), attitudes (i.e. to cybercrime, risk, self-protection) and technological characteristics (i.e. technology skill set, technology used, day to day technology brands, internet) of the victims to build a complete profile.

Following this information the common characteristics of the victims can be identified. For example, it may become apparent that the majority of victims were of the age 25 to 30, if this was the case preventative methods could be targeted to this age group such as educating them on how to stay safe online. Or, it may be the case that the majority of victims were found to be in full time employment and earning over £50,000 or were start-up companies with less than 50 employees. In each case preventative methods need to be, in part, targeted to the victims found most at risk. With improvements to data collection statistical analysis can be performed to establish which characteristics of cybercrime victims are statistically significant. Researching the reasons why these characteristics have left people vulnerable to cyber-attacks will aid the development of preventative methods and evidence base on cybercrime.

The term cybercrime encompasses a multitude of offences, of which require different skill sets and knowledge (i.e. phishing, skimming, espionage). The offenders themselves may be fuelled by the same or very different motives and intentions depending on the type of cyber offence. Are the victims therefore the same or different dependent on the type of cybercrime? The motives and intentions of one cybercriminal to the next when committing

the same type of crime may be quite different but there may be some similarities in the victims they target, which therefore need to be investigated. Can the characteristics of a person or business be predictive of the attacks they are at risk of encountering?

Summary

Analysis of both cybercriminals and their victims will develop our understanding of cybercrime and aid preventative measures to tackle cybercriminal activity. The analysis of cybercriminals needs to expand on current research, looking in depth at their characteristics and criminal patterns whilst comparing the findings to traditional offenders. Following the production of longitudinal data on cybercriminals, the advancement of cybercrime research will be further amplified by the introduction of criminal career research; enhancing our evidence base.

In understanding cybercrime we do not have to stop at offender research, analysis into victim profiles will also provide valuable information and widen the evidence base. Establishing the reasons why some individuals become victim to cybercrime when others do not will aid the development of preventative methods to combat cybercrime. Extensive and innovative research is needed to further the fight against such crime.

Conclusion

The workshop and our research highlighted the difficulty with regard to identifying statistical information for computer enabled crime, unlike computer dependent crime which has a reasonably robust set of criteria for assessment as embodied by legislation such as the Computer Misuse Act. Computer enabled crime embodies the transformation of conventional crime (covered by current legislation) by digital technology.

This poses a significant issue for data collectors and the legislature in where to draw the line. This report has attempted to define new approaches that may help in this regard through the focus on the impact of the technology in such cases via an assessment against metrics such as force multiplier and entry barrier. However, this discussion leads to the intriguing possibility of an alternative to developing new legislation to handle computer enabled crime. Instead new sentencing guidelines could be developed for existing criminal offences where technology is seen as an aggravating or reducing factor in the commissioning of the crime. This provides the court service greater flexibility in sentencing, but also enables the courts to remain dynamic enough to keep up to date with the evolving use of technology in the commission of crimes.

Although currently the form of cybercrime data is patchy at best, the full depth of its wealth has not yet been explored with the potential to provide some very valuable and intriguing information. This exploration should take place before the development of new data begins to depict any potentially rich information which can be built on or incorporated. In addition, knowing the limitations of existing data will aid the development of future data sources. The data source taxonomy proposed will provide researchers with the tools to classify and evaluate the quality of existing data for their research restricting the risk of unaccounted bias. To move forward in this space it was debated theoretically that the production and provision of a standardised data frame was needed to develop reliable and valid datasets whilst facilitating optimal sample sizes and comparable datasets. Following the development in quantity and quality of publically available data advancements in statistical analysis can be made. Through the use of statistical techniques factors associated to cybercriminals can be identified along with recurring themes found within victims. Such information can be used in addition to current research to not only help in the detection of cyber criminals but also target resources and preventative measures.

While this report has focused on the types of data that could be collected and how this could be achieved, what it has not discussed is why this has not already been achieved. In

2004 the chairman UK all Party Internet Group, Derek Wyatt, highlighted the need to obtain pertinent information on the extent of the cybercrime problem. In 2007 David Wall summarised the state of play with regards to the available data in that there were "encouraging signs" as US and UK national crime and justice surveys were including questions on cybercrime victimisation. However, the results published by industry need to be scrutinised in order to avoid the potentially self-fulfilling fear, uncertainty and doubt markets model for security products. Wall also highlighted a serious lack of information on offenders and their motivation, accompanied with a significant amount of victim under-reporting. In 2011 the UK government set up its strategy to promote the UK as the best place in the world to conduct digital business. This included the following objective:

"By the end of 2011, build a single reporting system for citizens and small businesses to report cybercrime so that action can be taken and law enforcement agencies can establish the extent of cybercrime (including how it affects individuals and the economy)."

During the discussion at the workshop with the key stakeholders, it became clear that gathering detailed information on cybercrime statistics has begun to move on (i.e. Action Fraud) but is still in need of significant advancement. Arguably the government's goal of becoming the number one cyber secure business location cannot be validated without appropriate evidence. In nine years our empirical knowledge of offending and victimisations has only just begun to develop. So why has the will not be translated into action? What are the inhibitors here? And yet despite a lack of knowledge the criminal justice system has not collapsed. This indicates to the authors' two potential results:

1. That the criminal justice system has adapted and self-organised to accommodate these digital activities within its existing processes and statues, or,
2. Criminal activity is going undetected and unprosecuted.

It is not clear to us which is the case. What is clear however, further research is required before we can begin to answer these questions.

Writing Team

About the Authors

Claire Hargreaves: Current PhD student in Applied Social Statistics at Lancaster University

Claire Hargreaves is completing a PhD in Applied Social Statistics at Lancaster University. Her primary research is in quantitative criminology which focuses on the criminal careers of offenders. During her PhD she has worked for the Home Office on a 6 month internship and undertaken collaborative work with Statistics Norway.

Dr Daniel Prince: Associate Director of Partnership at Security Lancaster

Daniel Prince is an associate director for Security Lancaster, managing business partnerships and enterprise. Prior to this he was the course director for the multi-disciplinary MSc in Cyber Security teaching penetration testing, digital forensics and information security risk management. Daniel holds a PhD in Computer Science and has worked on projects with numerous large technology companies such as Cisco and Microsoft.

About the Contributors

Prof Brian Francis: Professor in the Department of Mathematics and Statistics at Lancaster University

Prof Brian Francis is a social statistician with interests in preference models, latent class methods for longitudinal data and categorical data problems. Applications are crucial to his work and he is primarily research is in quantitative criminology -both methodological and substantive problems - focusing on the criminal careers of offenders and social and developmental change. He is also interested in analysing administrative and survey data and investigating issues related to violence and serious offending, including sexual offending and homicide. More broadly he is working on child developmental studies in psychology and in foetal development.

David Cook: Associate Solicitor at Pannone

David is an Associate Solicitor, Solicitor Advocate, in the Regulatory team at Pannone Solicitors and specialises in cyber crime and data security matters. David has extensive experience across a broad range of offences but in particular specialises in cases of a complex nature. He has a particular interest in crimes and civil wrongdoing committed using computers or through the internet. David is widely regarded as an expert in this area and regularly appears in the national and local media.

Pannone Solicitors is firmly established as one of the UK's leading law firms. They advise on all aspects of law with teams specialising in corporate services, dispute resolution and regulatory, family, personal and financial and injury and negligence.

With over 600 staff, including over 300 specialist lawyers, Pannone Solicitors serve a wide variety of clients across the UK and internationally, including private individuals, public sector organisations and business clients ranging from established SMEs through to mature multi-nationals. Their lawyers, many of whom are leaders in their chosen fields, share a passion for winning and achieving the best possible outcome for every client.

Serious Organised Crime Agency (SOCA)

SOCA tackles serious organised crime that affects the UK and our citizens. This includes Class A drugs, people smuggling and human trafficking, major gun crime, fraud, computer crime and money laundering.

References

Farrington, D.P. (1992). 'Criminal Career Research in the United Kingdom,' *British Journal of Criminology,* Vol.32, No.4.

Francis, B., Soothill, K. and Ackerley, E. (2004). 'Multiple Cohort Data, Delinquent Generations, and Criminal Careers,' *Journal of Contemporary Criminal Justice*, Vol. 20, No.2, May, 103-126.

Holt, T. J. (2010). Exploring strategies for qualitative criminological and criminal justice inquiry using on-line data. *Journal of Criminal Justice Education*, Vol.21, 300-321.

Holt, T. J., Soles, J., and Leslie, L. (2008). Characterizing malware writers and computer attackers in their own words. Paper presented at the 3rd International Conference on Information Warfare and Security, April 24-25, in Omaha, Nebraska.

Holt, T.J., Strumsky, D., Smirnova, O. & Kilger, M. (2012). Examining the Social Networks of Malware Writers and Hackers. *International Journal of Cyber Criminology,* Vol.6, No.1, 891-903.

Macloed, J.F., Grove, P.G. and Farrington, D.P. (2012). *Explaining Criminal Careers: Implications for Justice Policy,* Oxford: Oxford University Press.

Soothill, K., Fitzpatrick, C. and Francis, B. (2009). *Understanding Criminal Careers*, Devon: Willon Publishing.

Wall, D.S. (2007/10).Policing Cybercrimes: Situating the Public Police in Networks of Security within Cyberspace (Revised May 2010), *Police Practice & Research: An International Journal,* Vol.8, No.2, 183-205.

Security Futures' mission is to is create a space where we could develop innovative techniques to think about the future, techniques that draw together the insight and expertise of researchers working across different disciplines. In this collaborative space, researchers and other partner organisations have the freedom to explore questions about security and technology. But also to formulate the questions that we might need to start asking about the emerging trends in technology, society and security. A space where we can bring together people working on the cutting edges of technology, social, legal and political disciplines to ask questions about the world we live in. A space where we might begin to imagine new horizons and start to see the problems that

Security Lancaster is a university wide research centre on security and protection sciences. It delivers research and education that innovates and creatively challenges the way that individuals, organisations and societies secure and protect themselves. This is achieved via engagement and collaboration with organisations from a range of sectors along with governments. The centres approach delivers the very best use-inspired and pure research alongside cutting edge education that delivers real impact and social change.

This work was funded by Security Lancaster via the auspices of Lancaster University's Faculty of Science and Technology.

Science and Technology Business Partnerships and Enterprise

As well as working with a range of external partners, ICT and Security form part of a wider theme based team across Science and Technology at Lancaster who offer expertise in:

- Advanced Manufacturing
- Energy
- Environment
- Health & Human Development
- Quantum Technologies
- Mathematics and Statistics

Working in Partnership

Across the themes we form collaborative partnerships around these 5 key areas:

- Collaborative Research and Consultancy
- Training and Education
- Co-location and Secondment
- Student Placements
- Product Development and IPR

For more information and video highlights of the workshop please visit

http://www.security-centre.lancs.ac.uk/cyber-criminal